

Sensitivity Equations Provide More Robust Gradients and Faster Computation of the FOCE Approximation to the Population Likelihood

Joachim Almquist^{1,2}, Jacob Leander^{1,3,*}, and Mats Jirstrand¹

¹Fraunhofer-Chalmers Centre, Göteborg, ²Department of Chemical and Biological Engineering, Chalmers University of Technology, Göteborg

³Department of Mathematical Sciences, Chalmers University of Technology, Göteborg, *Current affiliation AstraZeneca R&D, Mölndal

Background

The first order conditional estimation (FOCE) method [1] is still one of the parameter estimation workhorses for nonlinear mixed effects (NLME) modeling used in population pharmacokinetics and pharmacodynamics [2]. We propose an novel implementation of the FOCE and FOCEI methods where instead of obtaining the gradients needed for the two levels of quasi-Newton optimizations from the standard finite difference approximation, gradients are computed using so called sensitivity equations [3].

The Approximate Population Likelihood

The state-space model for a single individual is described by a system of ordinary differential equations and a corresponding set of measurement equations

$$\begin{aligned} \frac{dx_i(t)}{dt} &= f(x_i(t), t, Z_i(t), \theta, \eta_i) & y_{ij} &= h(x_{ij}, t_{ij}, Z_i(t_{ij}), \theta, \eta_i) + e_{ij} \\ x_i(t_0) &= x_{0i}(Z_i(t_0), \theta, \eta_i) & e_{ij} &\in N(0, R_{ij}(x_{ij}, t_{ij}, Z_i(t_{ij}), \theta, \eta_i)) \\ & & \hat{y}_{ij} &= E[y_{ij}] \end{aligned}$$

where indices i and j denote individuals and observations, respectively. Furthermore, θ are fixed effects parameters, $Z_i(t_j)$ are covariates, $\eta_i \sim N(0, \Omega)$ are random effect parameters, and R_{ij} are measurement error covariance matrices.

Given a set of experimental observations, d_{ij} , for the individuals $i = 1, \dots, N$ at the time points t_{ij} , where $j = 1, \dots, n_i$, we define the residuals $\epsilon_{ij} = d_{ij} - \hat{y}_{ij}$

The approximate log-likelihood function is obtained using the Laplacian approximation, which involves a second order Taylor expansion wrt η_i of l_i around points η_i^* that maximizes the individual l_i .

$$\log L(\theta) \approx \log L_F(\theta) = \sum_{i=1}^N \left(l_i(\eta_i^*) - \frac{1}{2} \log \det \left[\frac{-H_i(\eta_i^*)}{2\pi} \right] \right)$$

where

$$l_i = -\frac{1}{2} \sum_{j=1}^{n_i} \left(\epsilon_{ij}^T R_{ij}^{-1} \epsilon_{ij} + \log \det(2\pi R_{ij}) \right) - \frac{1}{2} \eta_i^T \Omega^{-1} \eta_i - \frac{1}{2} \log \det(2\pi \Omega)$$

The Outer Optimization Problem

The outer optimization problem consists of finding the θ that maximizes the log-likelihood. The m^{th} component of the gradient of the log-likelihood wrt θ

$$\frac{d \log L_F}{d \theta_m} = \sum_{i=1}^N \left(\frac{dl_i(\eta_i^*)}{d \theta_m} - \frac{1}{2} \text{tr} \left[H_i^{-1}(\eta_i^*) \frac{dH_i(\eta_i^*)}{d \theta_m} \right] \right)$$

where the total derivatives of l_i and H_i wrt θ can be expressed in terms of solutions to the sensitivity differential equations, e.g.,

$$\begin{aligned} \frac{dl_i(\eta_i^*)}{d \theta_m} &= \frac{dl_i(\eta_i)}{d \theta_m} \Big|_{\eta_i=\eta_i^*(\theta)} = \left[-\frac{1}{2} \sum_{j=1}^{n_i} \left(2\epsilon_{ij}^T R_{ij}^{-1} \frac{d\epsilon_{ij}}{d \theta_m} - \epsilon_{ij}^T R_{ij}^{-1} \frac{dR_{ij}}{d \theta_m} R_{ij}^{-1} \epsilon_{ij} \right. \right. \\ &\quad \left. \left. + \text{tr} \left[R_{ij}^{-1} \frac{dR_{ij}}{d \theta_m} \right] \right) + \frac{1}{2} \eta_i^T \Omega^{-1} \frac{d\Omega}{d \theta_m} \Omega^{-1} \eta_i - \frac{1}{2} \text{tr} \left[\Omega^{-1} \frac{d\Omega}{d \theta_m} \right] \right] \Big|_{\eta_i=\eta_i^*(\theta)} \\ \frac{d\epsilon_{ij}^*}{d \theta_m} &= \frac{d\epsilon_{ij}}{d \theta_m} \Big|_{\eta_i=\eta_i^*(\theta)} + \frac{d\epsilon_{ij}}{d \eta_i} \Big|_{\eta_i=\eta_i^*(\theta)} \frac{d\eta_i^*}{d \theta_m} \quad \text{where} \quad \frac{d\epsilon_{ij}}{d \theta_m} = \frac{d(d_{ij} - \hat{y}_{ij})}{d \theta_m} = - \left(\frac{\partial h}{\partial \theta_m} + \frac{\partial h}{\partial x_{ij}} \frac{dx_{ij}}{d \theta_m} \right) \end{aligned}$$

The sensitivity differential equations wrt θ_m

$$\frac{d}{dt} \left(\frac{dx_i}{d \theta_m} \right) = \frac{\partial f}{\partial \theta_m} + \frac{\partial f}{\partial x_i} \left(\frac{dx_i}{d \theta_m} \right) \quad \left(\frac{dx_i}{d \theta_m} \right)(t_0) = \frac{\partial x_{0i}}{\partial \theta_m}$$

How to find $\frac{d\eta_i^*}{d \theta}$?

$$\frac{dl_i}{d \eta_i} \Big|_{\eta_i=\eta_i^*(\theta)} = 0 \Rightarrow \frac{d}{d \theta} \left(\frac{dl_i}{d \eta_i} \Big|_{\eta_i=\eta_i^*(\theta)} \right) = 0 \Rightarrow \frac{d\eta_i^*}{d \theta} = - \left(\frac{d^2 l_i}{d \eta_i^2} \Big|_{\eta_i=\eta_i^*(\theta)} \right)^{-1} \frac{d^2 l_i}{d \eta_i d \theta} \Big|_{\eta_i=\eta_i^*(\theta)}$$

Second order sensitivities are also required: $\frac{d^2 x_i}{d \eta_{ik} d \theta_m}$ and $\frac{d^2 x_i}{d \eta_{ik} d \eta_{il}}$.

The Inner Optimization Problem

The inner optimization problem consists of finding the η_i that maximizes the individual l_i (for a given θ). Gradient based optimization methods need accurate gradients. The k^{th} component of the gradient of the log-likelihood wrt η_i

$$\frac{dl_i}{d \eta_{ik}} = -\frac{1}{2} \sum_{j=1}^{n_i} \left(2\epsilon_{ij}^T R_{ij}^{-1} \frac{d\epsilon_{ij}}{d \eta_{ik}} - \epsilon_{ij}^T R_{ij}^{-1} \frac{dR_{ij}}{d \eta_{ik}} R_{ij}^{-1} \epsilon_{ij} + \text{tr} \left[R_{ij}^{-1} \frac{dR_{ij}}{d \eta_{ik}} \right] \right) - \eta_i^T \Omega^{-1} \frac{d\eta_i}{d \eta_{ik}}$$

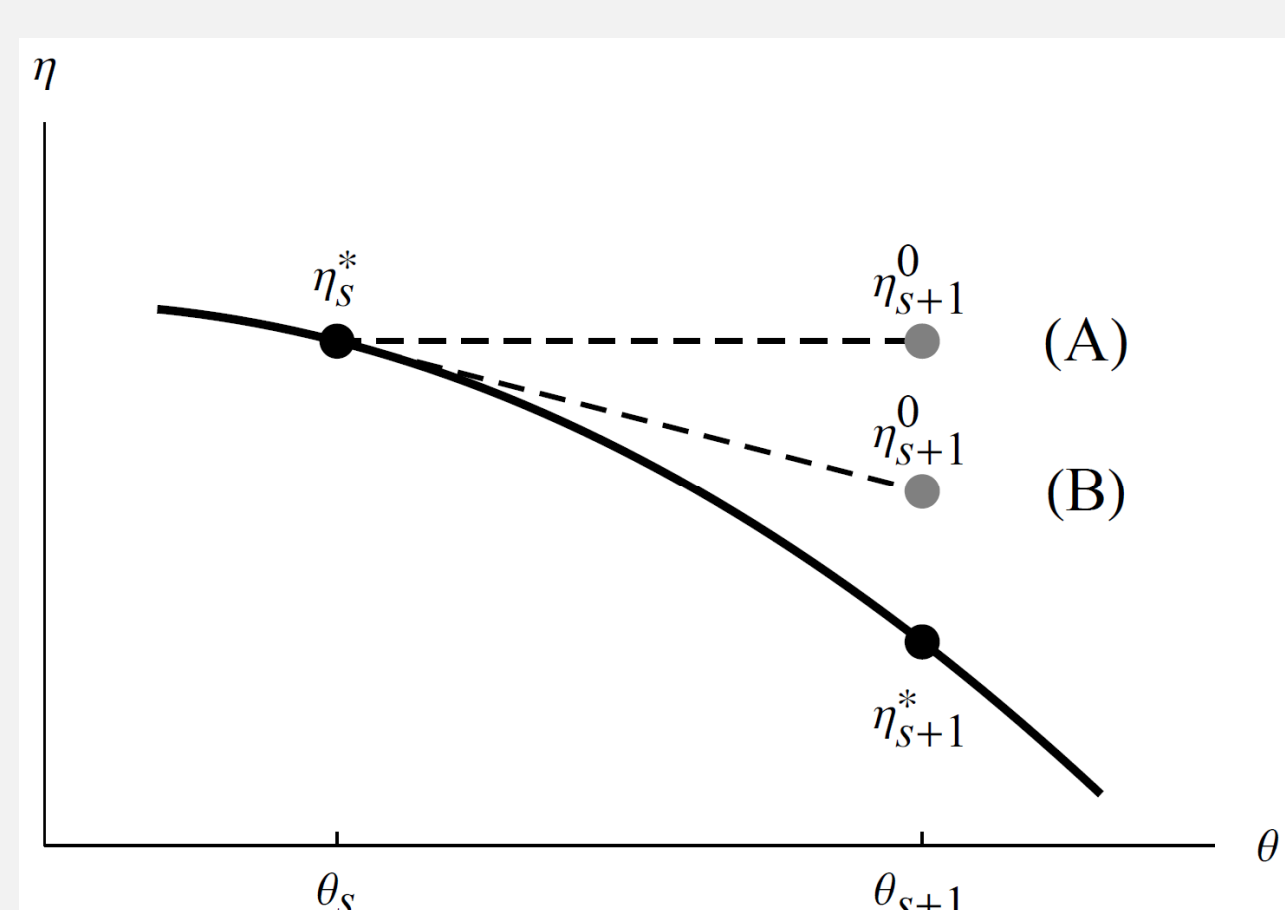
where

$$\frac{d\epsilon_{ij}}{d \eta_{ik}} = \frac{d(d_{ij} - \hat{y}_{ij})}{d \eta_{ik}} = - \left(\frac{\partial h}{\partial \eta_{ik}} + \frac{\partial h}{\partial x_{ij}} \frac{dx_{ij}}{d \eta_{ik}} \right) \quad \text{and} \quad \frac{dR_{ij}}{d \eta_{ik}} = \frac{\partial R_{ij}}{\partial \eta_{ik}} + \frac{\partial R_{ij}}{\partial x_{ij}} \frac{dx_{ij}}{d \eta_{ik}}$$

The sensitivity differential equations wrt η_{ik}

$$\frac{d}{dt} \left(\frac{dx_i}{d \eta_{ik}} \right) = \frac{\partial f}{\partial \eta_{ik}} + \frac{\partial f}{\partial x_i} \left(\frac{dx_i}{d \eta_{ik}} \right) \quad \left(\frac{dx_i}{d \eta_{ik}} \right)(t_0) = \frac{\partial x_{0i}}{\partial \eta_{ik}}$$

Starting Values for Random Parameters

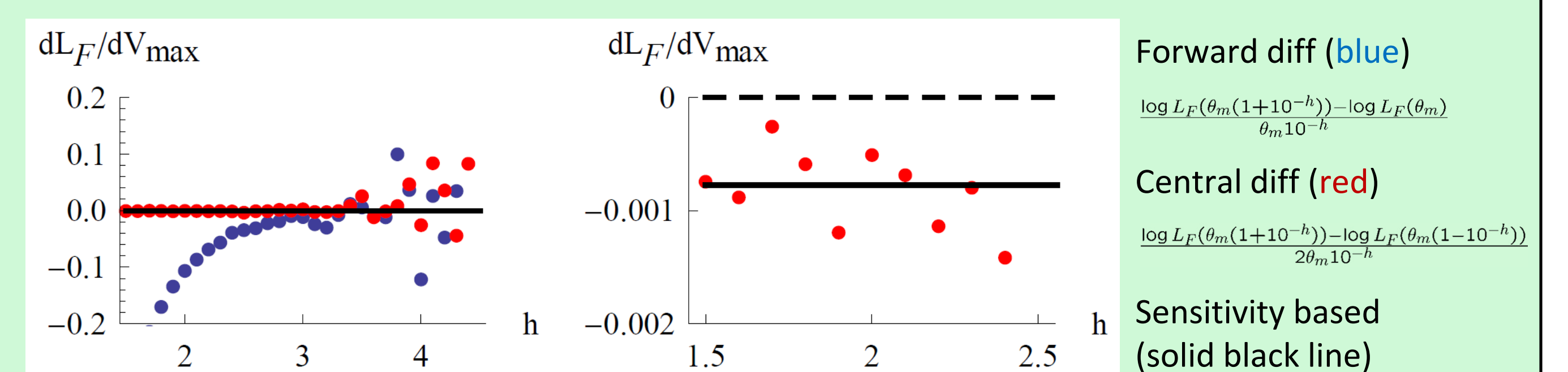


Using that $\eta_i^* = \eta_i^*(\theta)$ is a function of θ and that we have $\frac{d\eta_i^*}{d \theta}$ gives improved starting values of the inner optimization problem

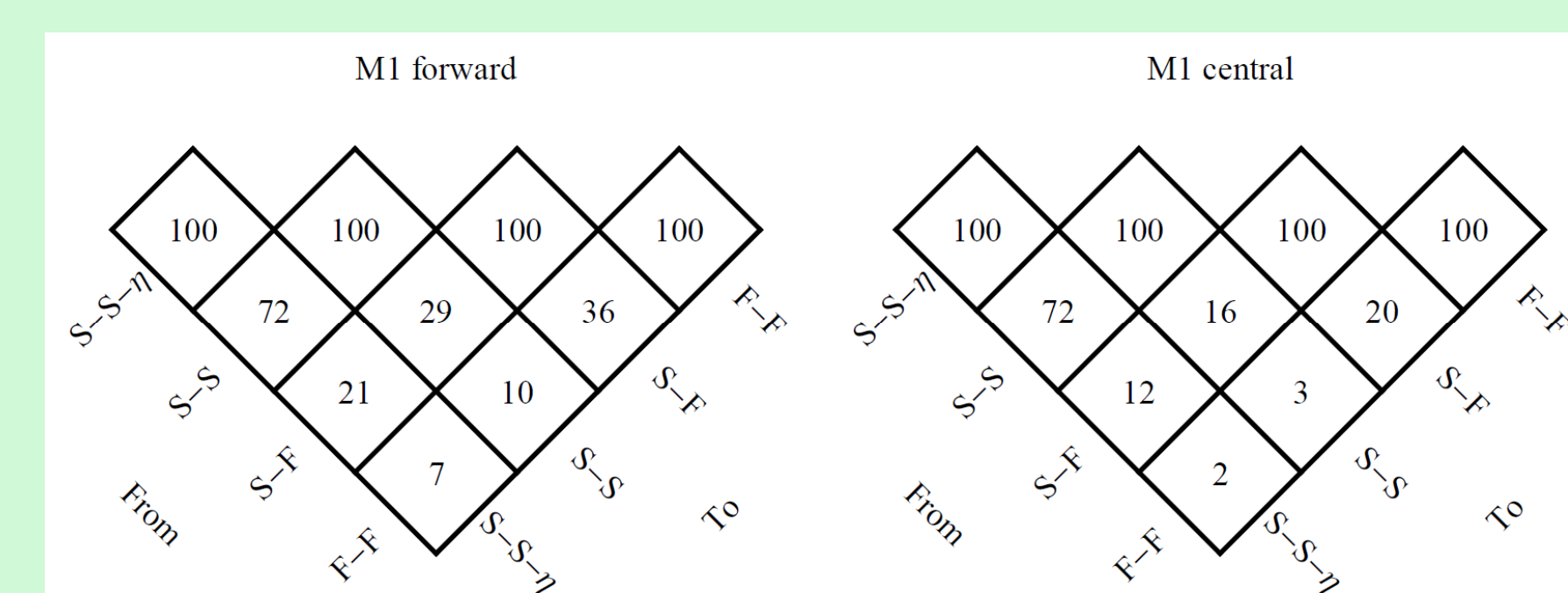
$$\eta_{s+1}^0 = \eta_s^* + \frac{d\eta_s^*}{d \theta} (\theta_{s+1} - \theta_s)$$

Precision, Accuracy, and Performance

Two different levels of magnification of an element of the log-likelihood gradient as a function of the finite difference step, h .



Benchmarking – relative estimation times



Model M1: 2-compartment, nonlinear elimination

S-F- η : Sensitivities (inner), Finite differences (outer), improved η starting values

Example: F-F (central diff) to S-S- η gives 50-fold decreased computational time

Highlights

- Robust computation of gradients
- Methodology applies to both individual and population log-likelihoods
- Improves computational speed compared to finite differences

References

- [1] Wang Y. Derivation of various NONMEM estimation methods. J of Pharmacokin Pharmacodyn (2007) 34(5): 575-593.
- [2] Johansson ÅM, Ueckert S, Plan EL, Hooker AC, Karlsson MO. Evaluation of bias, precision, robustness and runtime for estimation methods in NONMEM 7. J of Pharmacokin Pharmacodyn (2014) 41(3):223-238.
- [3] Almquist J, Leander J, Jirstrand M. Using sensitivity equations for computing gradients of the FOCE and FOCEI approximations to the population likelihood. J of Pharmacokin Pharmacodyn (2015) 42(3):191-209.

Acknowledgements

This project has in part been supported by the Swedish Foundation for Strategic Research, which is gratefully acknowledged.